

Q76360

(List of Cited Literature)

1. Japanese Unexamined Patent Application Publication 2002-049501
2. Japanese Unexamined Patent Application Publication H10-326199

(Reason 1)

(Claims) 1-22

(Cited literature) 1 and 2

(Remarks)

Cited Literature 1 describes a duplexed fault-tolerant system which performs identical operations, wherein a collation device collates I/O access requests to a first bus from the CPUs of information processing devices, and if a mismatch is detected, said CPUs are notified by means of an interrupt, and having received the interrupt, the CPUs are made to perform I/O access and the timing of the I/O access is adjusted to restart correct duplexed processing.

Cited Literature 2 describes an invention for a control method for a dual system comprising a counter indicating the progress of a program on the system in question and a register which stores a count value indicating the progress of a program on the other system, forwarded from the other system, whereby, in case of an interrupt based on an external interrupt signal, the count values of the two systems are compared, and in case of mismatch, the program execution of the system with the greater count value is stopped and placed in a standby state, and the program of the system with the smaller count value is allowed to proceed until the count values match, with interrupt being allowed when the count values of the two systems match.

Adapting the invention described in Cited Literature 2 to the invention described in Cited Literature 1 is a matter which could be easily conceived of by a person skilled in the art. Thus, in the invention described in Cited Literature 1, a person skilled in the art could easily conceive of an arrangement whereby, instead of the means which causes I/O access to be carried out and adjusts the I/O access timing, the CPUs are notified of a mismatch detection interrupt and the program execution of one of the information processing devices is put on standby until the count values indicating the program execution status of the information processing devices match, i.e., an arrangement whereby each processor is notified of the mismatch detection interrupt and delay adjustment is performed to match up the command execution status between the computing modules.

Therefore, it is found that the invention according to Claims 1 through 22 of the present application could have been easily invented by a person skilled in the art based on the inventions described in Cited Literature 1 and 2.

Record of Prior Art Literature Search Results

- | | |
|------------------------|---|
| • Fields searched | IPC G06F 11/18 |
| • Prior art literature | Japanese Unexamined Patent Application Publication H04-247531 |

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-049501

(43)Date of publication of application : 15.02.2002

(51)Int.Cl.

G06F 11/18
G06F 12/16
G06F 15/177

(21)Application number : 2000-237364

(71)Applicant : NIPPON TELEGR & TELEPH CORP
<NTT>

(22)Date of filing : 04.08.2000

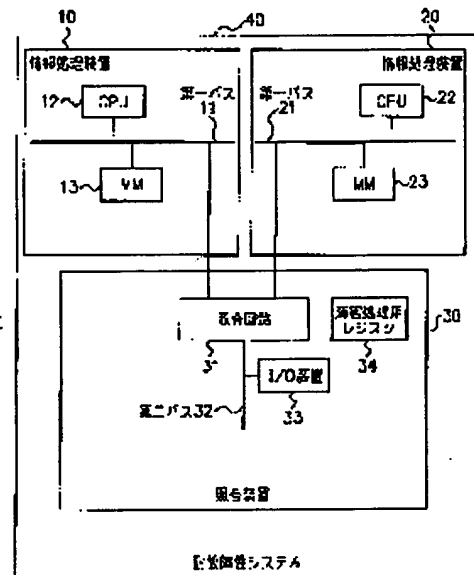
(72)Inventor : MISHIMA TAKESHI
MASUDA ETSUO

(54) FAULT-TOLERANT SYSTEM AND ITS FAULT DEMARCATING METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a fault-tolerant system and its fault demarcating method which can shorten the interruption time of service as much as possible by demarcating an information processor which has got out of order in a short time in dual systems.

SOLUTION: A collating circuit 31 collates to each other requests from information processors 10 and 20 which have CPUs 12, 22 and MM's 13, 23 respectively, where 1st buses 11 and 21 are connected to an I/O device 33 in connection to a 2nd bus 32. The circuit 31 judges that neither of the information processors 10 and 20 has got out of order when they do not match each other. When they do not match each other, the CPUs 11 and 21 are informed of that and a new access request is made on the basis of preliminarily prepared data; and those data are compared with data held in a register 34 for fault processing to decide which of the information processor has got out of order.



LEGAL STATUS

[Date of request for examination]

07.01.2002

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2002-49501
(P2002-49501A)

(43) 公開日 平成14年2月15日 (2002.2.15)

| (51) Int.Cl. ⁷ | 識別記号 | F I | テーマコード* (参考) |
|---------------------------|-------|---------------|-------------------|
| G 0 6 F 11/18 | 3 1 0 | G 0 6 F 11/18 | 3 1 0 C 5 B 0 1 8 |
| 12/16 | 3 1 0 | 12/16 | 3 1 0 L 5 B 0 3 4 |
| 15/177 | 6 7 8 | 15/177 | 6 7 8 A 5 B 0 4 5 |

審査請求 未請求 請求項の数 2 O L (全 8 頁)

(21) 出願番号 特願2000-237364 (P2000-237364)

(22) 出願日 平成12年8月4日 (2000.8.4)

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区大手町二丁目3番1号

(72) 発明者 三島 健

東京都千代田区大手町二丁目3番1号 日

本電信電話株式会社内

(72) 発明者 増田 悦夫

東京都千代田区大手町二丁目3番1号 日

本電信電話株式会社内

(74) 代理人 100069981

弁理士 吉田 精孝

Fターム (参考) 5B018 GA05 HA05 QA01

5B034 AA02 CC01

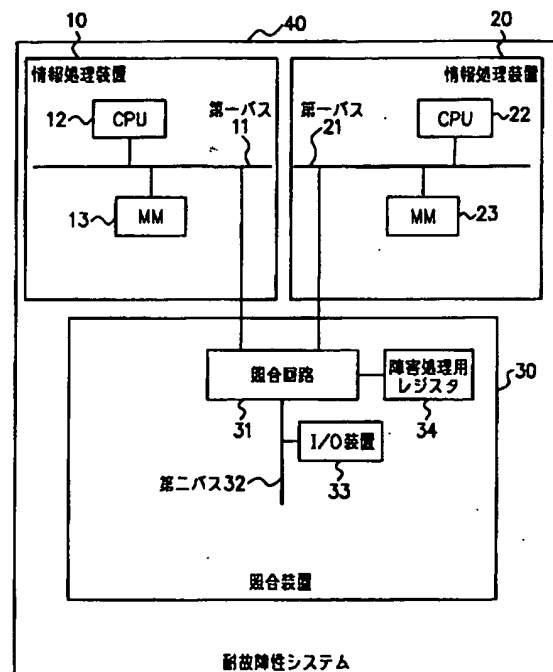
5B045 JJ03 JJ07 JJ23 JJ48

(54) 【発明の名称】 耐故障性システム及びその故障切り分け方法

(57) 【要約】

【課題】 二重化システムにおいて、故障した情報処理装置を短時間で切り分け、サービスの中断時間をできるだけ短くし得る耐故障性システム及びその故障切り分け方法を提供すること。

【解決手段】 それぞれCPU 12、22、MM 13、23を有し、照合装置30の照合回路31に第一バス11、21が接続された情報処理装置10、20からの、第二バス32に接続されたI/O装置33へのアクセス要求を照合回路31により照合し、一致すれば情報処理装置10、20に故障はないと判断し、不一致であればCPU 11、21に通知して予め決めておいたデータによるアクセス要求を発生させ、これらを障害処理用レジスタ34に保持しておいたデータと比較することにより、故障した情報処理装置を判別する。



【特許請求の範囲】

【請求項1】 それぞれ中央処理装置を有する二つの情報処理装置と、各情報処理装置を接続するとともに情報処理装置から出される要求の同一性を照合し故障を検出する照合手段とを有する耐故障性システムにおいて、前記照合手段は、

前記要求の照合の結果が不一致であった場合、各情報処理装置に通知する手段と、

各情報処理装置から情報を受け取り、該情報と照合手段自身が保持している情報とを合わせて三つの情報で多数決を取り、どちらの情報処理装置が故障であるかを判断する手段とにより構成したことを特徴とする耐故障性システム。

【請求項2】 それぞれ中央処理装置を有する二つの情報処理装置と、各情報処理装置を接続するとともに情報処理装置から出される要求の同一性を照合し故障を検出する照合手段とを有する耐故障性システムにおいて、前記照合手段は、前記要求の照合の結果が不一致であった場合、各情報処理装置に通知し、各情報処理装置から情報を受け取り、該情報と照合手段自身が保持している情報とを合わせて三つの情報で多数決を取り、どちらの情報処理装置が故障であるかを判断することを特徴とする耐故障性システムの故障切り分け方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、24時間連続運転が要求される通信網のノードシステム、企業内のサーバ等に利用される耐故障性システム及びその故障切り分け方法に関するものである。

【0002】

【従来の技術】図9は従来のこの種の耐故障性システムの一例、ここでは三重化システムを示すものである。

【0003】図9に示すように、この耐故障性システムでは同じスペックの三つの情報処理装置1010、1020、1030を備えている。各情報処理装置1010、1020、1030は、それぞれ中央処理装置（以下、CPU（Central Processing Unit）と言う）1011、1021、1031、主記憶装置1012、1022、1032、I/O装置1013、1023、1033及びそれらを接続するバス1014、1024、1034等から構成されている。

【0004】情報処理装置1010、1020、1030のバス1014、1024、1034には、照合回路1040が接続されている。照合回路1040は、バス1014、1024、1034のそれぞれに流れる情報を引き込み、三つの情報の多数決を取る。

【0005】三つの情報が一致している場合、故障はないと判断し、そのままサービスを継続する。三つの情報が不一致の場合、照合回路1040は異なる情報を出力した情報処理装置を故障と判断する。故障と判断された

情報処理装置は、耐故障性システム1050から切り離され、正常な二つの情報処理装置でサービスは継続される。

【0006】また、図10は従来のこの種の耐故障性システムの他の例、ここでは二重化システムを示すものである。

【0007】図10に示すように、この耐故障性システムでは同じスペックの二つの情報処理装置1060、1070を備えている。各情報処理装置1060、1070は、それぞれ中央処理装置（以下、CPUと言う）1061、1071、主記憶装置1062、1072、I/O装置1063、1073及びそれらを接続するバス1064、1074等から構成されている。

【0008】情報処理装置1060、1070のバス1064、1074には、照合回路1080が接続されている。照合回路1080は、バス1064、1074のそれぞれに流れる情報を引き込み、二つの情報を比較する。

【0009】二つの情報が一致している場合、故障はないと判断し、そのままサービスを継続する。二つの情報が不一致の場合、照合回路1080は、いずれかの情報処理装置に異常が発生したと判断し、その旨を（図示しない割り込み通知の線にて）両方の情報処理装置1060、1070へ通知する。そして、サービスは中断され、ヒストリ情報解析や状態情報解析、レジスタの内容の比較、レジスタ、メモリ、I/O装置等のアクセスの正常性確認等、様々な試験から成る障害処理が行われ、異常の有無を判定し、照合回路1080へ通知される。照合回路1080は、その情報から故障した情報処理装置を耐故障性システム1090から切り離し、正常な情報処理装置でサービスを再開する。

【0010】

【発明が解決しようとする課題】上述したように、従来の三重化による耐故障性システムにおける、故障した情報処理装置を切り分けは、三つの情報処理装置で全く同一の処理を行い、各情報処理装置のある情報で多数決を行うことで実現し、また、従来の二重化による耐故障性システムにおける、故障した情報処理装置を切り分けは、二つの情報処理装置で全く同一の処理を行い、各情報処理装置のある情報の比較を行い、不一致が生じた場合には、どちらの情報処理装置が故障であるかを特定するために、ヒストリ情報解析や状態情報解析、レジスタの内容の比較、レジスタ、メモリ、I/O装置等のアクセスの正常性確認等、様々な試験を行うことで実現していた。

【0011】従って、従来の耐故障性システムについては、以下のような点が問題となっていた。

（1）三重化システムにした場合、コストが高くなる。

（2）二重化システムにした場合、故障した情報処理装置を切り分けるために多くの時間がかかり、長時間サー

ビスが中断される。

【0012】本発明は、上記事情を鑑みてなされたものであり、その目的とするところは、上記問題点を解決し、二重化システムにおいて、故障した情報処理装置を短時間で切り分け、サービスの中断時間をできるだけ短くし得る耐故障性システム及びその故障切り分け方法を提供することにある。

【0013】

【課題を解決するための手段】上記目的を達成するために、請求項1の発明では、それぞれ中央処理装置を有する二つの情報処理装置と、各情報処理装置を接続するとともに情報処理装置から出される要求の同一性を照合し故障を検出する照合手段とを有する耐故障性システムにおいて、前記照合手段は、前記要求の照合の結果が不一致であった場合、各情報処理装置に通知する手段と、各情報処理装置から情報を受け取り、該情報と照合手段自身が保持している情報とを合わせて三つの情報で多数決を取り、どちらの情報処理装置が故障であるかを判断する手段とにより構成したことを特徴とする耐故障性システムを提案する。

【0014】本発明によれば、情報処理装置からもらう二つの情報と照合手段が持っている情報とを合わせて三つの情報で多数決を取ることにより、短時間で故障の情報処理装置を判別することができ、サービスの中断を最小限の時間に抑えることができる。

【0015】また、請求項2の発明では、それぞれ中央処理装置を有する二つの情報処理装置と、各情報処理装置を接続するとともに情報処理装置から出される要求の同一性を照合し故障を検出する照合手段とを有する耐故障性システムにおいて、前記照合手段は、前記要求の照合の結果が不一致であった場合、各情報処理装置に通知し、各情報処理装置から情報を受け取り、該情報と照合手段自身が保持している情報とを合わせて三つの情報で多数決を取り、どちらの情報処理装置が故障であるかを判断することを特徴とする耐故障性システムの故障切り分け方法を提案する。

【0016】本発明によれば、情報処理装置からもらう二つの情報と照合手段が持っている情報とを合わせて三つの情報で多数決を取ることにより、短時間で故障の情報処理装置を判別することができ、サービスの中断を最小限の時間に抑えることができる。

【0017】

【発明の実施の形態】

【0018】

【第1の実施の形態】以下、本発明の耐故障性システムについて図面を参照して説明する。

【0019】図1は本発明の耐故障性システムの第1の実施の形態を示すものである。

【0020】この耐故障性システムは、図示するように、同一の処理を行う情報処理装置10及び20と、各

情報処理装置10及び20を接続する照合装置30とを備えている。

【0021】各情報処理装置10、20は、それぞれ第一バス11、21と、第一バス11、21に接続した中央処理装置（以下、CPUと言う）12、22と、主記憶装置（以下、MM（Main Memory）と言う）13、23とを備える。

【0022】照合装置30は、照合回路31と、該照合回路31に接続されている第二バス32とを備えており、I/O装置33は該第二バス32に接続されている。照合回路31は、障害処理用レジスタ34を備えている。

【0023】また、情報処理装置10、20は、それぞれ障害処理ソフトウェアを備えている。これらの障害処理ソフトウェアは、それぞれ障害切り分けに使うために、予め決めておいた同じデータ、例えば「xxxx」を持っており、耐故障性システム40の立ち上げ時に予めそのデータ「xxxx」を障害処理用レジスタ34に保持しておく。

【0024】第二バス32にI/O装置33を有する照合回路31は第一バス11、21に接続されている。

【0025】この耐故障性システム40では、情報処理装置10、20からI/O装置33へのアクセスを比較することで、情報処理装置10、20の故障を検出する。照合回路31は、比較の結果、一致だった場合のみI/O装置33へのアクセスを実行するように制御することによって、情報処理装置10、20の故障を、I/O装置33や他のシステムへ波及しないように工夫している。

【0026】図2を参照して本耐故障性システム40の動作について、説明する。

【0027】【ステップSA1】CPU12は、第一バス11のバス権を獲得した後、I/O装置33へI/Oアクセス要求を出す。CPU22は、第一バス21のバス権を獲得した後、I/O装置33へI/Oアクセス要求を出す。この時点では、照合回路31は、I/O装置33へアクセスするためのリード要求を第二バス32へ出さない。

【0028】なお、一定時間内にもう一方のCPU12または22からの照合対象のI/Oアクセス要求が来なかった場合、不一致が発生したと判断し、その旨を割り込みにて各情報処理装置10、20へ通知する。不一致後の処理は、後述の障害処理ソフトウェアが行う。

【0029】【ステップSA2】照合回路31によるCPU12、22のI/Oアクセス情報の比較の結果、一致の場合、両情報処理装置10、20は故障は無いので、照合回路31は、I/O装置33へアクセスするための要求を第二バス32へ出し、I/Oアクセスは正常終了する。

【0030】照合回路31によるCPU12、22のI

／Oアクセス情報の比較の結果、不一致の場合、照合回路31は割り込みにより照合不一致である旨をCPU12、22へ通知する。I／Oアクセスは中断し、CPU12、22は第一バス11、21のバス権を解放する。

【0031】[ステップSA3] 割り込みを受けたCPU12、22では、それぞれ障害処理ソフトウェアを走らせる。障害処理ソフトウェアは、予め決めておいたデータ「xxxx」をライトデータとしてI／Oライト命令を出す。

【0032】[ステップSA4] 照合回路31は、I／Oライト命令のデータと障害処理用レジスタ34内のデータ「xxxx」を比較する。もし、I／Oライト命令のデータが「xxxx」である場合、その命令を出した情報処理装置は正常であると判断する。もし、I／Oライト命令のデータが「xxxx」と異なる場合、あるいはI／Oライトが正常に行われなかった場合、その情報処理装置は故障と判断する。

【0033】[ステップSA5] 照合回路31は、故障を持っている情報処理装置を切り離し、正常な情報処理装置だけでサービスを再開する。両方とも一致した場合、間欠故障として、二重化のままサービスを再開する。

【0034】図3にこの処理フローを示す。

【0035】以上のように、この耐故障性システム40では、CPUからのI／O装置へのアクセスを照合回路31が仲介し、一致の場合は正常なI／O装置の処理を行う。不一致の場合は障害処理ソフトウェアが予め決めておいたデータを使ってI／Oアクセスし、そのI／Oデータと障害処理用レジスタ34が持っているデータとを合わせ、三つのデータで多数決を取ることによって故障した情報処理装置を判別する。短時間のサービスの中断後、正常系でサービスを継続できる。

【0036】

【第2の実施の形態】図4は本発明の耐故障性システムの第2の実施の形態を示すものである。

【0037】この耐故障性システムは、図示するように、同一の処理を行う情報処理装置210及び220と、各情報処理装置210及び220を接続する照合装置230とを備えている。

【0038】各情報処理装置210、220は、それぞれバス211、221と、バス211、221に接続した中央処理装置（以下、CPUと言う）212、222と、主記憶装置（以下、MMと言う）213、223と、I／O装置214、224とを備える。

【0039】照合装置230は、照合回路231を備え、また、照合回路231は障害処理用レジスタ232を備えている。

【0040】また、情報処理装置210、220は、それぞれ障害処理ソフトウェアを備えている。これらの障害処理ソフトウェアは、それぞれ障害切り分けに使うた

めに、予め決めておいた同じデータ、例えば「xxxx」を持っており、耐故障性システム240の立ち上げ時に予めそのデータ「xxxx」を障害処理用レジスタ231に保持しておく。

【0041】照合回路231はバス211、221に接続している。第1の実施の形態とは異なり、照合回路231はバス信号を止めることなく、単にバス211、221上に流れる情報をスヌープしている。

【0042】図5を参照して本耐故障性システム240の動作について、説明する。

【0043】[ステップSC1] CPU212は、バス211のバス権を獲得した後、I／O装置214へI／Oアクセス要求を出す。CPU222は、バス221のバス権を獲得した後、I／O装置224へI／Oアクセス要求を出す。照合回路231は、バス211、221に流れる情報をスヌープする。

【0044】なお、一定時間内にもう一方のCPU212または222からの照合対象のI／Oアクセス要求が来なかった場合、不一致が発生したと判断し、その旨を割り込みにて各情報処理装置210、220へ通知する。不一致後の処理は、後述の障害処理ソフトウェアが行う。

【0045】[ステップSC2] 照合回路231によるCPU212、222のI／Oアクセス情報の比較の結果、一致の場合、両情報処理装置210、220は故障は無いので、何も行わない。

【0046】照合回路231によるCPU212、222のI／Oアクセス情報の比較の結果、不一致の場合、照合回路231は割り込みにより照合不一致である旨をCPU212、222へ通知する。

【0047】[ステップSC3] 割り込みを受けたCPU212、222では、それぞれ障害処理ソフトウェアを走らせる。障害処理ソフトウェアは、予め決めておいたデータ「xxxx」をライトデータとして照合回路231へ転送（照合回路231へライト）する。

【0048】[ステップSC4] 照合回路231は、転送されたデータと障害処理用レジスタ232内のデータ「xxxx」を比較する。もし、転送されたデータが「xxxx」である場合、その命令を出した情報処理装置は正常であると判断する。もし、転送されたデータが「xxxx」と異なる場合、あるいは照合回路231へデータ転送が正常に行われなかった場合、その情報処理装置は故障と判断する。

【0049】[ステップSC5] 照合回路231は、故障を持っている情報処理装置を切り離し、正常な情報処理装置だけでサービスを再開する。両方とも一致した場合には、間欠故障として、二重化のままサービスを再開する。

【0050】図6にこの処理フローを示す。

【0051】以上のように、この耐故障性システム24

0では、バス上に流れる情報を常にスヌープし、情報の比較を行って故障の発生を監視する。比較の結果、不一致の場合は障害処理ソフトウェアが予め決めておいたデータを照合回路231へ転送し、障害処理用レジスタ232が持っているデータと合わせて三つのデータで多数決を取ることで、故障した情報処理装置を判別する。短時間のサービスの中断後、正常系でサービスを継続できる。

【0052】

【実施の形態のバリエーション】なお、上記実施の形態は例示的なものであり、本発明はこれに限定されるものではない。本発明の範囲は特許請求の範囲によって示されており、この特許請求の範囲の意味の中に入る全ての変形例は本発明に含まれるものである。

【0053】例えば、第1の実施の形態では、I/O装置33のみを照合装置30に配置しているが、例えば主記憶装置(MM)等も照合装置30側に配置し、メモリアクセスに関しても本発明による制御を行っても良い。

【0054】また、第1の実施の形態では情報処理装置の故障のみを対象としているが、図7に示すように、第2バス32を多重化することによりI/O装置の故障に対処できるように構成しても良い。

【0055】図7の例では、照合回路31に第二バス32と平行して第二バス32aを設けるとともに、該第二バス32aにI/O装置33aを接続している。即ち、第二バス32と第二バス32a、I/O装置33とI/O装置33aは互いに同一のものとする冗長構成である。このように構成することにより、さらに耐故障性が向上する。

【0056】さらに、本発明において、第1の実施の形態では、第一バス11、21と第二バス32の種類が一致していても、異なっても良い。

【0057】さらに、第1、第2の実施の形態では、照合装置30、230を1台のみ設けたが、例えば第1の実施の形態の応用として図8のように複数台(図8では2台)の照合装置30、30cを接続するようにしても良い。このように構成することにより、さらに耐故障性が向上する。

【0058】さらに、第1、第2の実施の形態では、障害処理用レジスタ34、232が照合回路31、231の外付けになっているが、照合回路31、231の内部

にあっても良い。

【0059】さらに、第1、第2の実施の形態では、障害処理用レジスタ34、232を使い、ハードワイヤードロジックで制御しているが、ファームウェアで制御しても良く、さらに障害処理用レジスタ34、232に格納するデータ「xxxx」をファームウェアのメモリ上に保持しても良い。

【0060】さらに、第1、第2の実施の形態を組み合わせても良い。

【0061】

【発明の効果】以上説明したように、本発明によれば、二重化による耐故障性システムにおいて、短時間に故障の情報処理装置を判別できる耐故障性システムが構築できる。従って、(1)安い耐故障性システムを構築できる、(2)サービスの中断時間を最小限にすることができる、といった効果が得られる。

【図面の簡単な説明】

【図1】本発明の耐故障性システムの第1の実施の形態を示す構成図

【図2】第1の実施の形態の動作を説明する図

【図3】第1の実施の形態の動作を説明する処理フロー

【図4】本発明の耐故障性システムの第2の実施の形態を示す構成図

【図5】第2の実施の形態の動作を説明する図

【図6】第2の実施の形態の動作を説明する処理フロー

【図7】本発明の耐故障性システムの他の実施の形態を示す構成図

【図8】本発明の耐故障性システムのさらに他の実施の形態を示す構成図

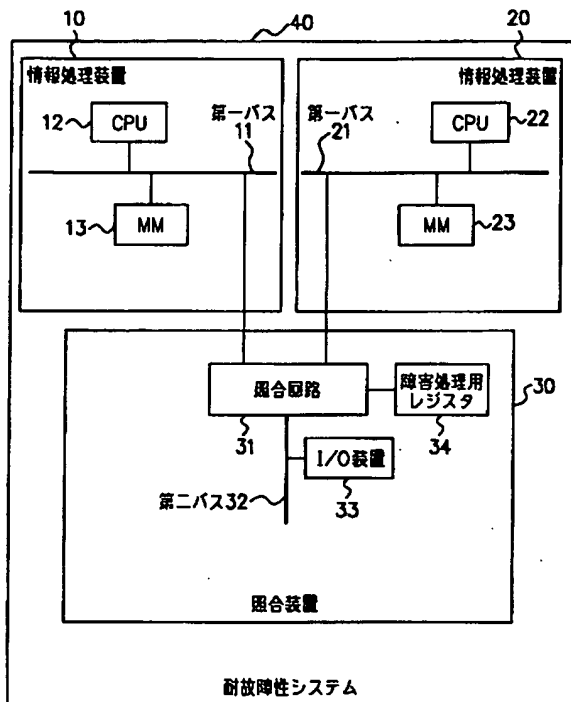
【図9】従来の耐故障性システムの一例を示す構成図

【図10】従来の耐故障性システムの他の例を示す構成図

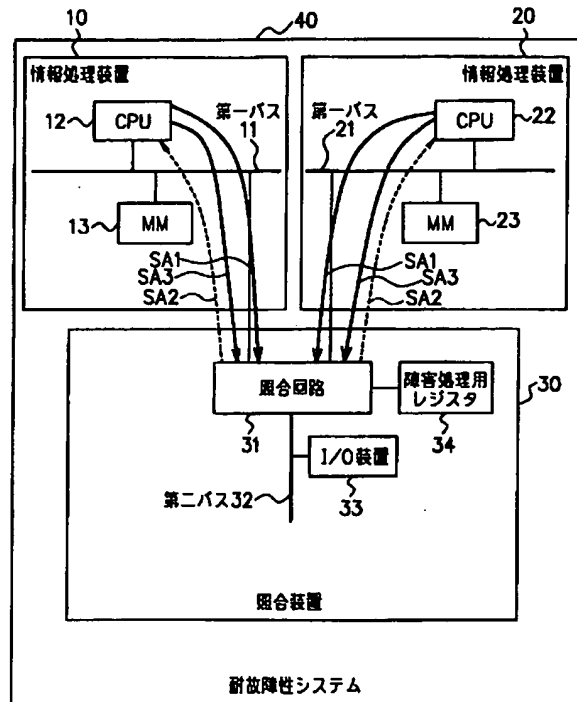
【符号の説明】

10、20、210、220：情報処理装置、11、21：第一バス、12、22、212、222：CPU、13、23、213、223：MM、30、30c、230：照合装置、31、231：照合回路、32、32a：第二バス、33、33a、214、224：I/O装置、34、232：障害処理用レジスタ、211、221：バス。

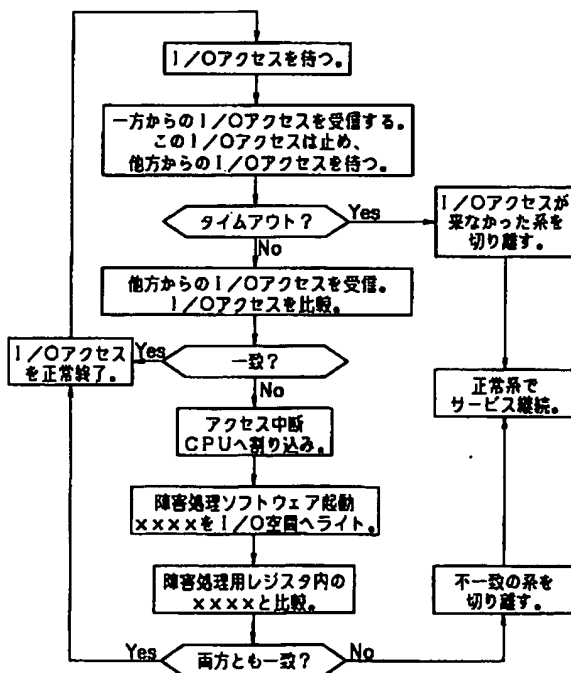
【図1】



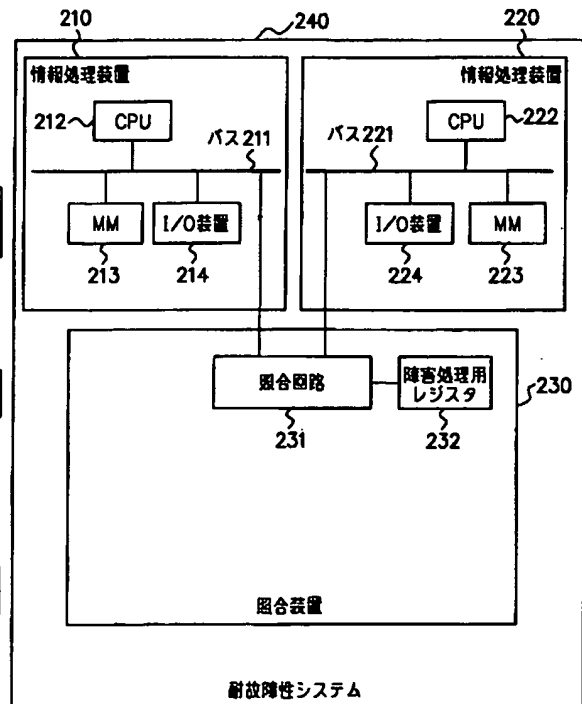
【図2】



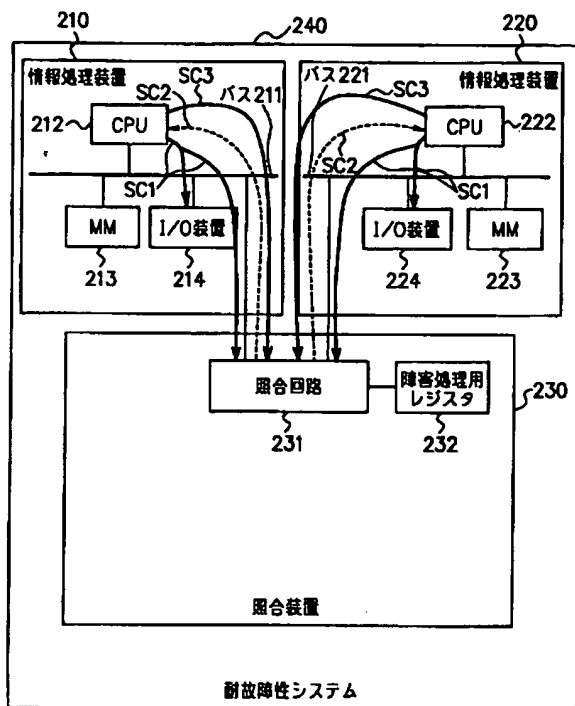
【図3】



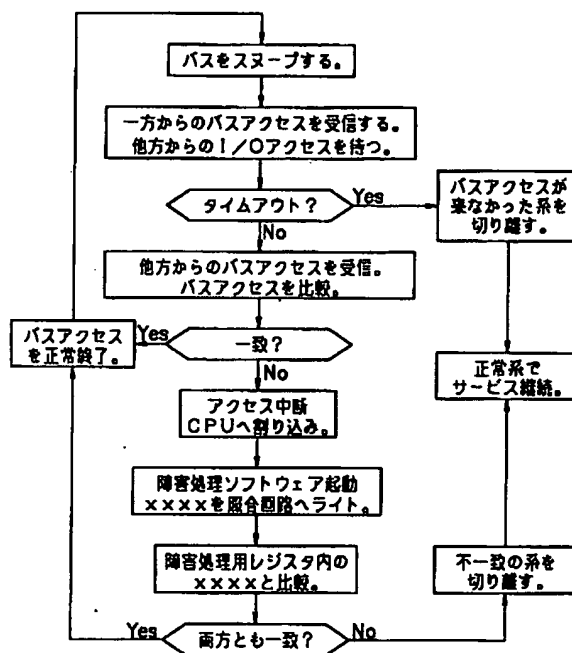
【図4】



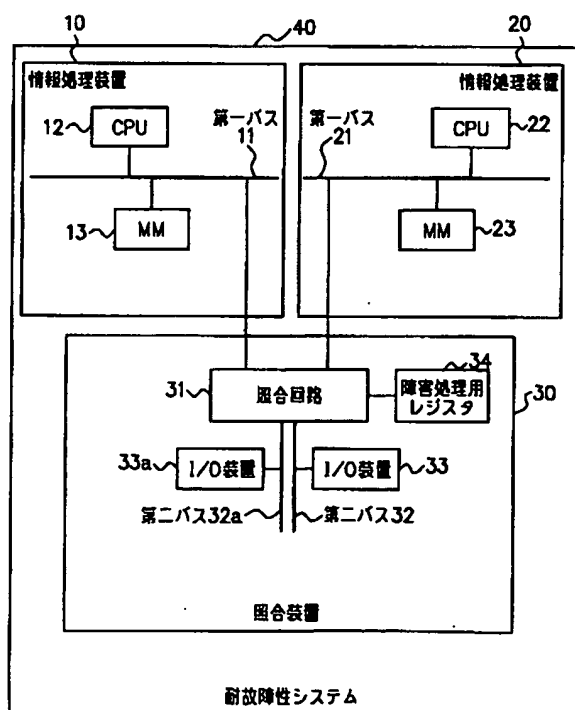
【図5】



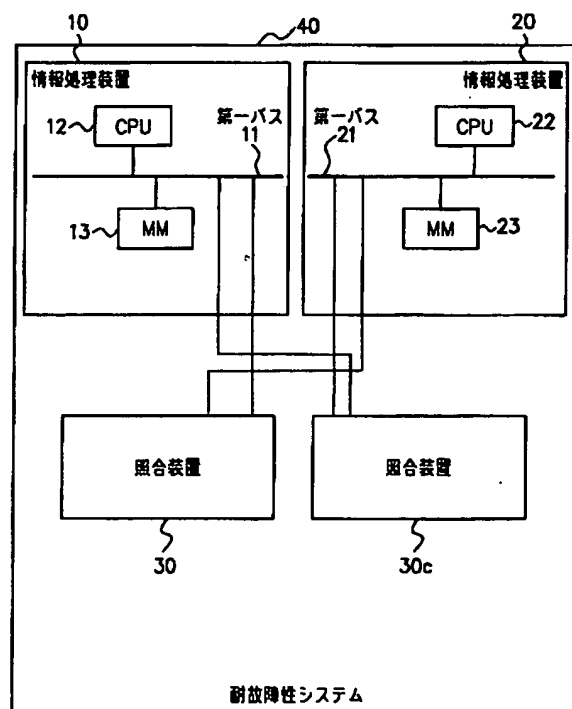
【図6】



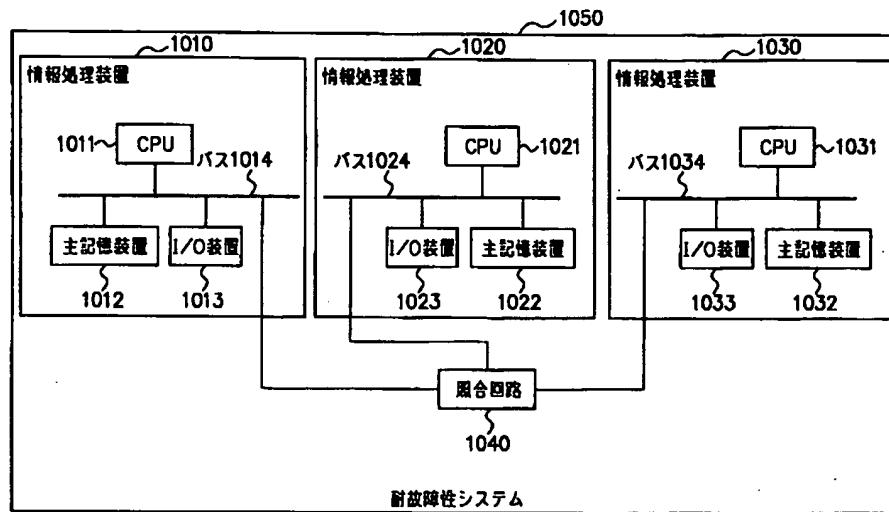
【図7】



【図8】



【図9】



【図10】

